JOURNAL OF COMBINATORIAL THEORY, Series A 63, 234–256 (1993)

The Number of Distinct Subset Sums of a Finite Set of Vectors

ERNEST BRICKELL*

Sandia National Laboratories, Albuquerque, New Mexico 87185

AND

MICHAEL SAKS†

Department of Mathematics, Rutgers University, New Brunswick, New Jersey 08903; and
Department of Computer Science and Engineering, University of California at San Diego, La Jolla, California 92093

Communicated by the Managing Editors

Received March 8, 1991

For a finite subset A of $\mathbf{R}^\rho$, $\sigma(A)$ denotes the set of vectors that can be obtained as a $\{0, 1\}$ linear combination of the vectors of A. Let $s_\rho(n)$ be the minimum cardinality of $\sigma(A)$ over all sets A of n nonnegative nonzero vectors that span $\mathbf{R}^\rho$. The main result is that for $n-\rho$ sufficiently large, $s_\rho(n)=2^{\rho-2}((n-\rho)^2+3(n-\rho)+4)$. This result is applied to answer a question of Erdős and Spencer concerning distinct sums of integers.

I. INTRODUCTION

The starting point for this paper is an elementary fact (observed by Ernst Strauss) concerning sums of positive integers. For a set J of integers (or vectors), let $\sum J$ denote the sum of the elements of J. Let $\sigma(J)$ denote the set of integers (or vectors) $\sum I$, where I ranges over all subsets of J. Obviously, if A has size n then $\sigma(J)$ has size at most 2^n, and this is achieved when J consists, for example, of distinct powers of 2. On the other hand, how small can $\sigma(J)$ be if J consists of n positive integers? Let $s(n)$ denote this minimum. For n an integer, let $[n]=\{1, 2, ..., n\}$.

PROPOSITION 1.1. *$s(n)=(n^2+n+2)/2$ which is achieved by the set $[n]$.*

* Supported in part by the U. S. Department of Energy under Contract DE-AC04-76DP00789; in part by NSF Grants DMS87 03541 and CCR-8922388.
† Supported in part by NSF Grants DMS87 03541 and CCR-8911388.

0097-3165/93 $5.00

In this paper, some analogous results for sets of vectors in Euclidean real space are obtained. The main question is: for positive integers ρ and n with $\rho \leqslant n$, what is the minimum cardinality of $\sigma(A)$ over all sets A of n nonzero nonnegative vectors having size n and rank ρ? (Here the rank of a set of vectors is the dimension of the subspace they span).

Let $\{\mathbf{e}_1, ..., \mathbf{e}_\rho\}$ be the standard basis of $\mathbf{R}^\rho$ and for any set J of reals, let $J^{(\rho)}$ denote the vector set $\mathbf{e}_1 J \cup \{\mathbf{e}_2, \mathbf{e}_3, ..., \mathbf{e}_\rho\}$, where $\mathbf{e}_1 J = \{j\mathbf{e}_1 : j \in J\}$. Note that $|\sigma(J^{(\rho)})| = 2^{\rho-1}|\sigma(J)|$. An obvious candidate for minimizing $\sigma(A)$ over sets A of nonzero nonnegative vectors having size n and rank ρ is the set $A = [n-\rho+1]^{(\rho)}$, for which $|\sigma(A)| = 2^{\rho-1}s(n-\rho+1)$. In general, however, this is not best possible: when $n = \rho+1$, this gives $|\sigma(A)| = 2^{\rho+1}$, while the set $B = \{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_\rho, \mathbf{e}_1 + \mathbf{e}_2\}$ has $|\sigma(B)| = 2^{\rho-2}(7)$.

The main result of this paper is that if $n-\rho$ is large enough then $[n-\rho+1]^{(\rho)}$ achieves the minimum. Actually, a slightly stronger statement is proved. A is an *antipode-free* set of vectors if $\mathbf{v} \in A$ implies $-\mathbf{v} \notin A$. Note that a set of nonzero vectors with nonnegative entries is antipode-free. Let $s_\rho(n)$ denote the minimum of $|\sigma(A)|$ over all sets A of n vectors that are antipode-free and span a space of dimension ρ.

THEOREM 1.2. *There exists a constant K such that for $n-\rho \geqslant K$,*

$$s_\rho(n) = 2^{\rho-1}s(n-\rho+1).$$

The condition that A be antipode-free is essential. Let $t_\rho(n)$ be the minimum of $|\sigma(A)|$ over all sets A of n nonzero vectors that span a space of dimension ρ and let $t(n) = t_1(n)$. For $\rho = 1$, we have

THEOREM 1.3. *Over sets J of nonzero real numbers, $|\sigma(j)|$ is minimized by the set $T(n) = \{1, 2, ..., \lceil n/2 \rceil\} \cup \{-1, -2, ..., -\lfloor n/2 \rfloor\}$; thus $t(n) = (n^2 + 2n + P(n))/4 + 1$ (where $P(n) = 1$ if n is odd and 0 otherwise).*

For $\rho > 1$ the set $T(n-\rho+1)^{(\rho)}$ shows that $t_\rho(n) \leqslant 2^{\rho-1}t(n-\rho+1)$. As with the antipode-free case, this is not always optimal but we have

THEOREM 1.4. *For $n-\rho \geqslant 5$, $t_\rho(n) = 2^{\rho-1}t(n-\rho+1)$.*

Another variant of this problem is obtained by allowing A to be a multiset. Let $u_\rho(n)$ be the minimum of $|\sigma(A)|$ over all multisets of n nonzero vectors than span a space of dimension ρ. In the 1-dimensional case, it is easy to see that $u_1(n) = n+1$; $|\sigma(A)|$ is minimized by taking n copies of the same element. In higher dimensions, the trivial construction is always optimal:

THEOREM 1.5. *For all positive integers $\rho \leqslant n$, $u_\rho(n) = 2^{\rho-1}(n-\rho+2)$.*

The investigation of $s_\rho(n)$ was motivated in part by a problem proposed by Erdős and Spencer about integer sums. The problem is most easily stated in terms of a game between two players MIN and MAX. In this game MIN selects a set A of a positive integers and MAX selects a set B of b positive integers, where A and B are disjoint. The objective of MIN is to minimize the size of $\sigma(A \cup B)$; MAX seeks to maximize the size of $\sigma(A \cup B)$. If MIN selects A first then MAX can select B to consist of distinct powers of 2 that are each greater than $\sum A$ and thus force the number of distinct sums to be $2^b|\sigma(A)|$, and MIN can do no better than to choose $A = [a]$. Thus if both players play optimally, the size of $\sigma(A \cup B)$ will be $2^b s(a)$.

The more interesting situation is if MAX must announce B first. Then MIN can choose $A = \{2i, 3i, ..., (a+1)i\}$, where i is the largest element of B, thus guaranteeing that the size of $\sigma(A \cup B)$ is at most $2^{b-1}s(a+1)$. However, in general this is not the best response by MIN. For instance, if $a = 1$, it is better for MIN to select his number to be the sum of any two numbers in B.

Let $\text{val}(a, b)$ denote the optimal value of the game when MAX goes first. Then $\text{val}(a, b) = \max_B \min_A |\sigma(A \cup B)|$, where A and B range over pairs of disjoint sets of positive integers with $|A| = a$ and $|B| = b$. As noted above, $\text{val}(a, b) \leqslant 2^{b-1}s(a+1)$. Erdős and Spencer asked whether this is close to the true answer. We relate $\text{val}(a, b)$ to the vector sum problem:

THEOREM 1.6. *For any nonnegative integers a and b,* $\text{val}(a, b) \geqslant s_b(a+b)$.

With Theorem 1.2 this implies

THEOREM 1.7. *There exists an integer K such that for* $a \geqslant K$, $\text{val}(a, b) = 2^{b-1}s(a+1)$,
(answering the question raised by Erdős and Spencer).

In fact, the proof of Theorem 1.6 shows that an optimal strategy for MAX is to choose any set $B = \{c_1, c_2, ..., c_b\}$, where for each index k, $c_{k-1}/c_k > 2n^{n-1}$, where $n = a + b$.

Section 2 contains definitions, notation, and facts about multisets of vectors. Section 3 contains a theorem about sums of multisets of real numbers that contains Proposition 1.1 and Theorem 1.3 as special cases. Section 4 gives some higher dimensional examples. Section 5 introduces the notion of criticality and states refinements of Theorem 1.2 (Theorem 5.3) and Theorem 1.4 (Theorem 5.2). Section 6 contains some properties of the function $|\sigma(A)|$, including the proof of Theorem 1.5. Section 7 contains some elementary arithmetic inequalities that are needed in the later sections. Section 8 contains the proof of Theorem 1.2, Section 9 has the proof of Theorem 1.4, and Section 10 has the proof of Theorem 1.6. Some open questions are mentioned in Section 11.

II. Preliminaries: Multisets of Vectors

A *multiset* M is a set X together with a function from X to the positive integers. The set X is called the *support* of M and is denoted $\underline{M}$. For $x \in \underline{M}$, $M(x)$ is called the *multiplicity* of x in M. We extend M to $x \notin \underline{M}$ by defining $M(x) = 0$. The *cardinality* of the multiset $|M|$ equals $\sum_{x \in \underline{M}} M(x)$. A multiset N is a *sub-multiset* of M, written $N \leqslant M$, if $N(x) \leqslant M(x)$ for all x. The sub-multiset *induced* on M by a set Y, denoted M_Y, is defined by $M_Y(x) = M(x)$ if $x \in Y$ and $M_Y(x) = 0$ otherwise. The difference of two multisets M and N, denoted $M - N$, is a multiset in which each element x gets multiplicity $\max\{M(x) - N(x), 0\}$. For an element y, $M - y$ denotes the multiset in which the multiplicity of y is reduced by 1 (unless it is already 0).

By abuse of notation, a multiset in which every element has multiplicity 1 is called a set and is treated as indistinguishable from its support. In particular, a subset of a multiset M is the same as a sub-multiset each of whose elements has multiplicity one.

All vector spaces in this paper are real and finite-dimensional. Vectors are usually written in boldface and $\mathbf{0}$ is the zero vector. Two vectors $\mathbf{v}$ and $\mathbf{w}$ are *antipodal* if $\mathbf{v} = -\mathbf{w}$. A set of vectors is *antipode-free* if it contains no antipodal pair of vectors (in particular it does not contain $\mathbf{0}$).

For a multiset M of vectors, $\langle M \rangle$ denotes the space spanned by M. Two vector multisets M and N are isomorphic if there is a vector space isomorphism ϕ between $\langle M \rangle$ and $\langle N \rangle$ such that $M(\mathbf{v}) = N(\phi(\mathbf{v}))$, for every vector $\mathbf{v}$ in $\langle M \rangle$. The *rank* of M, $\rho(M)$, is equal to the dimension of $\langle M \rangle$, and the *nullity* of M, $v(M)$, is equal to $|M| - \rho(M)$. If $\mathbf{W}$ is a subspace of $\langle M \rangle$, the sub-multiset $M_{\mathbf{W}}$ induced on M is called a *flat* of M. The set of flats is closed under intersection and this forms a lattice with $E \vee F$ equal to the multiset induced on M by $\langle E \cup F \rangle$. A flat of rank 1 consists of some element of M together with all multiples of it that belong to M. Note that every flat is a union of rank 1 flats. $\lambda(M)$ is the number of rank 1 flats and equal to the number of distinct "directions" determined by the elements of M.

A flat F is said to *cover* a flat E if F contains E and has rank 1 larger than E. A sequence of flats $\varnothing = F_0, F_1, \ldots, F_\rho = M$ in which F_j covers F_{j-1} is called a *flag* of M.

As usual, if $\mathbf{V}$ and $\mathbf{W}$ are vector spaces then $\mathbf{V}/\mathbf{W}$ denotes the quotient space of $\mathbf{V}$ with respect to $\mathbf{V} \cap \mathbf{W}$. Elements of this space are equivalence classes mod $\mathbf{W}$ which are typically denoted $\mathbf{v} + \mathbf{W}$, where $\mathbf{v}$ is an element of $\mathbf{V}$. Analogously, if M and N are multisets then M/N denotes the multiset defined on the vector space $\langle M \rangle / \langle N \rangle$ in which the multiplicity of the element $\mathbf{v} + \langle N \rangle$ is the *sum* of the multiplicities in M of the elements $\mathbf{w}$ that are equivalent to $\mathbf{v}$ mod $\langle N \rangle$. It is important to note that in general M/N

is not a set even if M is. For a vector $\mathbf{u}$, $M/\mathbf{u}$ is defined to be $M/\{\mathbf{u}\}$. It is easy to verify

PROPOSITION 2.1. *Let M be a vector multiset and N a multisubset. Then*

(i) $|M/N| = |M|$.

(ii) $\rho(M/N) = \rho(M) - \rho(N)$.

(iii) *If N is a flat of M then the sub-multiset of nonzero vectors of M/N has cardinality $|M| - |N|$.*

(iv) *If N is a flat of M then the flats of M/N are the sets F/N, where F ranges over flats of M that contain N.*

Let F be a proper flat of M. F is a *splitting flat* if $(M-F)/F$ is an antipode-free set. If any flat that covers F has size exactly $|F|+1$, then F is called a *plateau* of M.

PROPOSITION 2.2. *Let M be a vector multiset and F a flat of M.*

(i) *If F is a plateau of M, then $(M-F)/F$ is a set of pairwise linearly independent vectors and thus F is a splitting flat of M.*

(ii) *If M is an antipode-free set and $\rho(M-F) = \rho(M) - \rho(F)$ then F is a splitting flat.*

Proof. (i) By Proposition 2.1(iv), the rank 1 flats of $(M-F)/F$ are the sets $(E-F)/F$ where E is a flat that covers F. Hence, every rank 1 flat of $(M-F)/F$ has cardinality 1, and any two vectors are pairwise linearly independent.

(ii) If $\rho(M-F) = \rho(M) - \rho(F)$, then by elementary linear algebra $\langle F\rangle \cap \langle M-F\rangle = \langle \mathbf{0}\rangle$. Suppose, contrary to the proposition, that F is not a splitting flat. Then there are two vectors $\mathbf{v}$ and $\mathbf{w}$ belonging to $M-F$ such that either $\mathbf{v}-\mathbf{w}$ or $\mathbf{v}+\mathbf{w}$ belongs to $\langle F\rangle$. But they both belong to $\langle M-F\rangle$, which implies that one of them is $\mathbf{0}$. This contradicts the hypothesis that M is an antipode-free set. ∎

If $\mathbf{v}$ has positive multiplicity in M, then the *antipode class* of $\mathbf{v}$ in M is the sub-multiset induced by $\{\mathbf{v}, -\mathbf{v}\}$. The number of distinct nonzero antipode classes of M is denoted $\alpha(M)$. The *profile* of M, $\pi(M)$ is the sequence $(\pi_1, \pi_2, \ldots, \pi_{\alpha(M)})$ of sizes of the antipode classes in nonincreasing order. Two multisets M, N are *antipodally equivalent* if for any vector $\mathbf{v}$, $M(\mathbf{v}) + M(-\mathbf{v}) = N(\mathbf{v}) + N(-\mathbf{v})$. In particular two antipodally-equivalent multisets have the same profile.

The following result will be useful:

PROPOSITION 2.3. *Let A be a set of vectors.*

(i) *A has a rank one flat F such that $\alpha((A-F)/F) \geqslant (\lambda(A)-1)/3$.*

(ii) *If A is antipode-free and F is a flat of rank 1, then every antipode class of $(A-F)/F$ has size at most $\lambda(A)-1$.*

(iii) *If A is antipode-free and $|A| < 3\rho(A)/2$, then there is a rank one flat F of size 1 such that $(A-F)/F$ is an antipode-free set; i.e., $\alpha(A/F) = |A| - 1$.*

Proof. (i) Let $\lambda = \lambda(A)$ and let $X = \{\mathbf{v}_1, ..., \mathbf{v}_\lambda\}$ be a set of nonzero representatives from each dimension one flat of A. For $1 \leqslant k \leqslant \lambda$, let G_k be the graph on $X - \{\mathbf{v}_k\}$ with $\{\mathbf{v}_i, \mathbf{v}_j\}$ an edge of G_k if $\mathbf{v}_i + \langle \mathbf{v}_k \rangle$ and $\mathbf{v}_j + \langle \mathbf{v}_k \rangle$ are either equal or antipodal vectors in $A/\mathbf{v}_k$, i.e., if and only if one of the vectors $\mathbf{v}_i - \mathbf{v}_j$ or $\mathbf{v}_i + \mathbf{v}_j$ is a multiple of $\mathbf{v}_k$. Note that G_k is a disjoint union of complete graphs corresponding to the antipode classes of $A/\mathbf{v}_k$, and $\alpha(A/\mathbf{v}_k)$ is equal to the independence number of G_k. Each pair $\{\mathbf{v}_i, \mathbf{v}_j\}$ appears as an edge in at most two of the graphs G_k, and so the sum of the cardinalities of their edge sets is at most $\lambda(\lambda-1)$. Thus, for some index h, G_h has at most $\lambda - 1$ edges, which is less than or equal to the number of vertices. It follows from Turan's theorem [1], or by a simple induction, that any graph with v vertices and at most v edges has an independent set of size $v/3$. (The induction hypothesis is that every graph with v vertices and e edges has an independent set of size at least $(2v-e)/3$).

(ii) Let $E \neq F$ be a rank one flat of A. Then E/F is isomorphic to E and so all of the vectors in E are in distinct antipode classes of A/F, and thus any antipode class of $(A-F)/F$ contains at most one vector from each flat of $A-F$.

(iii) Let E be the largest flat of A that satisfies $\rho(E) \leqslant 2|E|/3$. Then E is not equal to A, by hypothesis. Let $\mathbf{u}$ be any vector in $A-E$, and let $F = \langle \mathbf{u} \rangle$. Then F has cardinality 1 since otherwise $E \vee F$ satisfies $\rho(E \vee F) \leqslant 1 + \rho(E) \leqslant 1 + 2|E|/3 \leqslant 2|E \vee F|/3$, contradicting the maximality of E. Suppose that there are two vectors $\mathbf{v}$ and $\mathbf{w}$ whose images mod F are either equal or antipodes. The flat G of A induced by $\langle \mathbf{u}, \mathbf{v}, \mathbf{w} \rangle$ has rank 2. Also, E contains at most one of $\mathbf{v}$ and $\mathbf{w}$ (since $\mathbf{u}$ is not in E). If E and G are disjoint then $\rho(E \vee G) \leqslant 2 + \rho(E) \leqslant 2 + 2|E|/3 \leqslant 2(|E| + 3)/3 \leqslant 2|E \vee G|/3$ and if not then $\rho(E \vee G) \leqslant 1 + \rho(E) \leqslant 1 + 2|E|/3 \leqslant 2(|E| + 2)/3 \leqslant 2|E \vee G|/3$, either one contradicting the maximality of E. Thus, every vector of $(A-F)/F$ is in a distinct antipode class. ∎

For a vector multiset M, let $\sigma(M)$ denote the set (not the multiset) of vectors $\sum_{v \in M} a_v \mathbf{v}$ where a_v is an integer between 0 and $M(v)$. Thus, if A is

a set then $\sigma(A)$ is the set of all sums of subsets of A. The reader can easily verify:

PROPOSITION 2.4. *Let M and N vector multisets. Then*

(i) *$\sigma(M/N)$ is equal to the support of $(\sigma(M-N))/N$.*

(ii) $|\sigma(M/N)| \leqslant |\sigma(M-N)|$.

The following result will be useful:

LEMMA 2.5. $|\sigma(M)| = |\sigma(N)|$ *if M and N are antipodally equivalent multisets.*

Proof. Let $d(\mathbf{v}) = N(\mathbf{v}) - M(\mathbf{v})$; then $d(-\mathbf{v}) = -d(\mathbf{v})$. Let A consist of one representative from each antipode class of M and let $D = \sum_{\mathbf{v} \in A} d(\mathbf{v})\mathbf{v}$. Then $\sigma(M)$ is equal to the set of vectors $\sum_{v \in A} a_{\mathbf{v}}\mathbf{v}$ where $a_{\mathbf{v}}$ is an integer between $-M(-\mathbf{v})$ and $M(\mathbf{v})$, and $\sigma(N)$ is equal to the set of vectors $\sum_{\mathbf{v} \in A} b_{\mathbf{v}}\mathbf{v}$, where $b_{\mathbf{v}}$ is an integer between $-N(-\mathbf{v}) = d(\mathbf{v}) - M(-\mathbf{v})$ and $N(\mathbf{v}) = d(\mathbf{v}) + M(\mathbf{v})$. Then the sum $S = \sum_{\mathbf{v} \in A} a_{\mathbf{v}}\mathbf{v}$ is in $\sigma(M)$ if and only if the sum $S + D = \sum_{\mathbf{v} \in A} (a_{\mathbf{v}} + d(\mathbf{v}))\mathbf{v}$ is in $\sigma(N)$. ∎

Recall from the introduction that $u_\rho(n)$, $t_\rho(n)$, and $s_\rho(n)$ are respectively the minimum of $|\sigma(M)|$ over multisets, sets and antipode-free sets of nonzero vectors of rank ρ and size n.

Let M be a multiset of vectors of rank ρ and cardinality n, and let $\mathbf{v}$ be any vector not in $\langle M \rangle$. If $N = M + \{\mathbf{v}\}$ then N is a multiset of rank $\rho + 1$, cardinality $n + 1$ and $|\sigma(N)| = 2|\sigma(M)|$. From this observation, we obtain:

LEMMA 2.6. *Let $\rho \leqslant n$ be integers greater than* 1. *Then*

(i) $s_\rho(n) \leqslant 2s_{\rho-1}(n-1)$.

(ii) $t_\rho(n) \leqslant 2t_{\rho-1}(n-1)$.

(iii) $u_\rho(n) \leqslant 2u_{\rho-1}(n-1)$.

III. SUMS OF REAL NUMBERS

Proposition 1.1 and Theorem 1.3 characterize the minimum of $|\sigma(A)|$ over sets of positive real numbers and over sets of arbitrary nonzero real numbers. These theorems can be deduced from

THEOREM 3.1. *Let M be a multiset of nonzero real numbers with profile $\pi = (\pi_1, \pi_2, ..., \pi_{\alpha(M)})$. Then $|\sigma(M)| \geqslant 1 + \sum_i (i\pi_i)$, and this is attained by the multiset $R(\pi)$ that assigns multiplicity π_i to each positive integer i.*

Before proving this theorem, we prove some corollaries. First of all, the same bound holds if M is a set of nonzero vectors.

COROLLARY 3.2. *Let M be a multiset of nonzero real vectors with profile $\pi = (\pi_1, \pi_2, ..., \pi_{\alpha(M)})$. Then $|\sigma(M)| \geq 1 + \sum_i (i\pi_i)$.*

Proof. By induction on $\rho(M)$. If $\rho(M) = 1$, the result is Theorem 3.1. Assume that $\rho(M) > 1$. It is enough to show that there is a vector $\mathbf{u}$ (not in M) such that $M/\langle \mathbf{u} \rangle$ has the same profile as M since $|\sigma(M/\langle \mathbf{u} \rangle)| \leq |\sigma(M)|$. Let $\mathbf{u}$ be any vector that is not a multiple of either $\mathbf{v} + \mathbf{w}$ or $\mathbf{v} - \mathbf{w}$ for any two vectors of M belonging to distinct antipode classes. Then the antipode classes of $M/\langle \mathbf{u} \rangle$ are the same as for M, as required. ∎

Specializing this corollary to the case of antipode-free sets and arbitrary sets of vectors we obtain

COROLLARY 3.3. *For any positive integers ρ and n with $\rho \leq n$,*

(i) $s_\rho(n) \geq (n^2 + n + 2)/2$.

(ii) $t_\rho(n) \geq (n^2 + 2n + P(n))/4 + 1$, *(where $P(n) = 0$ if n is even and 1 if n is odd).*

Proof. (i) Take all of the π_i to be 1 in Corollary 3.2.

(ii) The profile of an arbitrary set of n nonzero real numbers has all of the π_i equal to 1 or 2 and $\sum_i \pi_i = n$. Subject to this the minimum of $1 + \sum_i (i\pi_i)$ is $(n^2 + 2n + P(n))/4 + 1$ which is attained when $\pi_i = 2$ for $i \leq n/2$ and $\pi_{(n+1)/2} = 1$ if n is odd. ∎

Theorems 1.1 and 1.3 are obtained by taking $\rho = 1$ in the previous corollary.

Proof of Theorem 3.1. The second claim follows by noting that $\sigma(R(\pi))$ consists of all integers between 0 and $\sum_i (i\pi_i)$.

By Lemma 2.5, it suffices to prove the lower bound for the case that M consists entirely of positive numbers. Let $j_1 < j_2 \cdots < j_\alpha$ be the elements of M. We prove by induction on $|M|$ that $|\sigma(M)| \geq 1 + \sum_{i \leq \alpha} (i)\, M(j_i)$, which is at least $1 + \sum_i (i\pi_i)$. This is trivial for $|M| = 1$. Suppose $|M| > 1$ and let N be the sub-multiset obtained by reducing the multiplicity of j_α by 1. By induction $|\sigma(N)| \geq 1 + \sum_{i \leq \alpha - 1} (i)\, M(j_i) + (\alpha)(M(j_\alpha) - 1)$. Let T_α be the sum of all of the elements of N (with multiplicity) and for $i < \alpha$, let $T_i = T_\alpha - j_i$. Note that these are distinct sums belonging to $\sigma(N)$ and all are $> T_\alpha - j_\alpha$. Thus the α sums $T_i + j_\alpha$ are distinct for each i and are larger than anything in $\sigma(N)$, so $|\sigma(M)| \geq |\sigma(N)| + \alpha$, proving the desired result. ∎

IV. Some Examples in Higher Dimensions

The following examples show that the inequalities of Lemmas 2.6(i) and (ii) may be strict and thus $s_\rho(n)$ may be less than $2^{\rho-1}s(n-\rho+1)$ and $t_\rho(n)$ may be less that $2^{\rho-1}t(n-\rho+1)$.

Let $\{\mathbf{v}, \mathbf{w}, \mathbf{x}\}$ denote a basis of $\mathbf{R}^3$.

A. *Antipode-Free Sets*

Example 4.1. $\rho=2$, $n=3$. $A=\{\mathbf{v}, \mathbf{w}, \mathbf{v}+\mathbf{w}\}$. Then $|\sigma(A)|=7$, while $2s(n-1)=8$.

Example 4.2. $\rho=2$, $n=4$. $A=\{\mathbf{v}, \mathbf{w}, \mathbf{v}+\mathbf{w}, \mathbf{v}+2\mathbf{w}\}$. Then $|\sigma(A)|=12$, while $2s(n-1)=14$.

Example 4.3. $\rho=2$, $n=5$. $A=\{\mathbf{v}, \mathbf{w}, \mathbf{v}+\mathbf{w}, \mathbf{v}+2\mathbf{w}, \mathbf{v}+3\mathbf{w}\}$. Then $|\sigma(A)|=20$, while $2s(n-1)=22$.

B. *Unrestricted Sets*

Example 4.4. $\rho=2$, $n=4$. $A=\{\mathbf{v}, \mathbf{w}, -\mathbf{v}, -\mathbf{w}\}$. Then $|\sigma(A)|=9$ and $2t(n-1)=10$.

Example 4.5. $\rho=2$, $n=6$. $A=\{\mathbf{v}, \mathbf{w}, -\mathbf{v}, -\mathbf{w}, \mathbf{v}+\mathbf{w}, -\mathbf{v}-\mathbf{w}\}$. Then $|\sigma(A)|=19$ and $2t(n-1)=20$.

Example 4.6. $\rho=3$, $n=6$. $A=\{\mathbf{v}, \mathbf{w}, -\mathbf{v}, -\mathbf{w}, \mathbf{x}, -\mathbf{x}\}$. Then $|\sigma(A)|=27$, and $2t_2(n-1)=28$.

The last assertion in Example 4.6 depends on the (as yet unproven) fact that $t_2(5)=14$. This will follow from Theorem 5.4, but is also not hard to verify by case analysis.

V. Critical Indices

In this section, we introduce the notion of *criticality* which is used to formulate refinements of Theorems 1.2, 1.4, and 1.5.

A pair (ρ, v) of nonnegative integers with $\rho \geq 2$ is said to be an *s-critical index* if the inequality of Lemma 2.6(i) is strict; i.e., $s_\rho(\rho+v) < 2s_{\rho-1}(\rho+v-1)$. Define *t*-critical and *u*-critical indices similarly. Note that the functions s, t, and u are each completely determined by their values on their critical indices. Theorem 1.5 is equivalent to the following:

THEOREM 5.1. *There are no u-critical indices.*

On the other hand, Examples 4.4, 4.5, and 4.6 show that (2, 2), (2, 4), and (3, 3) are t-critical. In fact, as we shall see, these are the only t-critical indices. Define the function

$$\varepsilon(m) = \begin{cases} \frac{1}{2} & \text{if } m \in \{3, 5\} \\ \frac{1}{4} & \text{if } m = 4 \\ 0 & \text{otherwise,} \end{cases}$$

and let $t^*(m) = t(m) - \varepsilon(m)$. Finally, define

$$t_\rho^*(n) = \begin{cases} 2^{\rho-1}(t(n-\rho+1)) & \text{if } \rho = 1 \text{ or } (n = 5 \text{ and } \rho = 2) \\ 2^{\rho-1}(t^*(n-\rho+1)) & \text{otherwise.} \end{cases}$$

THEOREM 5.2. *For all ρ and n, $t_\rho(n) = t_\rho^*(n)$. Thus the only t-critical indices are* (2, 2), (2, 4), *and* (3, 3).

Examples 4.1, 4.2, and 4.3 show that (2, 1), (2, 2), and (2, 3) are s-critical. We have not been able to determine all of the s-critical indices, but we can establish:

THEOREM 5.3. *There are only finitely many s-critical indices.*

This implies Theorem 1.2, since we can take K in Theorem 1.2 to be greater than the maximum v over all s-critical indices (ρ, v).

The proofs of these results are presented in the next four sections.

VI. LOWER BOUNDS ON $|\sigma(M)|$

In this section we develop some techniques for obtaining lower bounds on $|\sigma(M)|$ for a specific multiset M, and prove Theorem 1.5. The main fact used is

LEMMA 6.1. *Let M and N be multisets of vectors such that $N \leqslant M$. Then*

$$|\sigma(M)| \geqslant |\sigma(M-N)| + |\sigma(M/N)|\,(|\sigma(N)| - 1).$$

To prove this lemma, we need a preliminary result. For two sets A and B of vectors let $A \oplus B = \{\mathbf{a} + \mathbf{b} \mid \mathbf{a} \in A, \mathbf{b} \in B\}$.

LEMMA 6.2. *For any two finite nonempty sets A and B of vectors in $\mathbf{R}^d$, $|A \oplus B| \geqslant |A| + |B| - 1$.*

Proof. Let $\mathbf{a}^* \in A$, $\mathbf{b}^* \in B$ be arbitrary and define $A^* = \{\mathbf{a} - \mathbf{a}^* \mid \mathbf{a} \in A\}$ and $B^* = \{\mathbf{b} - \mathbf{b}^* \mid \mathbf{b} \in B\}$. Then $|A^* \oplus B^*| = |A \oplus B|$ since $A^* \oplus B^* = \{\mathbf{c} - \mathbf{a}^* - \mathbf{b}^* \mid \mathbf{c} \in A \oplus B\}$, and so it is enough to prove the result for A^* and B^*. Note that $\mathbf{0}$ belongs to both A^* and B^*. For each $\mathbf{a} \in A^*$ let $\beta_{\mathbf{a}}$ be the greatest real number such that $\beta_{\mathbf{a}}\mathbf{a}$ belongs to B^* (since $\mathbf{0} \in B^*$, $\beta_{\mathbf{a}} \geqslant 0$). The vectors of the form $(\beta_{\mathbf{a}} + 1)\mathbf{a}$ are distinct and the set $C = \{(\beta_{\mathbf{a}} + 1)\mathbf{a} \mid \mathbf{a} \in A^* - 0\}$ is disjoint from B^* (by the choice of $\beta_{\mathbf{a}}$) and thus $|A^* \oplus B^*| \geqslant |B^* \cup C| = |A| + |B| - 1$. ∎

Proof of Lemma 6.1. Let $D_1, \ldots, D_h$ denote the partition of $\sigma(M - N)$ into equivalence classes mod $\langle N \rangle$. Now, $\sigma(M) = \sigma(M - N) \oplus \sigma(N) = \bigcup_{i \geqslant 1} (D_i \oplus \sigma(N))$ which is a disjoint union. Thus by Lemma 6.2, $|\sigma(M)| \geqslant \sum_{1 \leqslant i \leqslant h} (|D_i| + |\sigma(N)| - 1) = |\sigma(M - N)| + h(|\sigma(N)| - 1)$. By Proposition 2.4(i), $h = |\sigma(M/N)|$ and the lemma follows. ∎

An immediate consequence of Lemma 6.1 and Proposition 2.4(ii) is

COROLLARY 6.3. *Let M and N be multisets of vectors such that $N \leqslant M$. Then*

$$|\sigma(M)| \geqslant \sigma((M - N)/N)|(|\sigma(N)|).$$

COROLLARY 6.4. *Let M be a vector multiset of cardinality n and rank ρ and let $F_0, F_1, F_2, \ldots, F_\rho$ be a flag. Then $|\sigma(M)| \geqslant \prod_{1 \leqslant j \leqslant \rho} (|F_j - F_{j-1}| + 1)$.*

Proof. By Corollary 6.3, $|\sigma(F_k)| \geqslant |\sigma(F_{k-1})|\, |\sigma((F_k - F_{k-1})/F_{k-1})|$ for each k between 1 and ρ, and so $|\sigma(M)| = |\sigma(F_\rho)| = \prod_{1 \leqslant j \leqslant \rho} |\sigma((F_j - F_{j-1})/F_{j-1})|$. Since $(F_j - F_{j-1})/F_{j-1}$ consists of $|F_j - F_{j-1}|$ nonzero (though not necessarily distinct), vectors in dimension 1, Theorem 3.1 implies that $|\sigma((F_j - F_{j-1})/F_{j-1})| \geqslant |F_j - F_{j-1}| + 1$. ∎

The reader can verify

PROPOSITION 6.5. *The minimum of $\prod_{1 \leqslant j \leqslant d} a_j$ over all sequences $a_1, a_2, \ldots, a_d$ of positive integers that sum to m and are all at least c is $c^{d-1}(m - (d-1)c)$.*

Theorem 1.5 now follows from the previous two results.

Proof of Theorem 1.5. Let M be any multiset of cardinality n and rank ρ, let $F_0, F_1, F_2, \ldots, F_\rho$ be a flag, and $a_j = |F_j - F_{j-1}| + 1$. Then the a_j sum to $n + \rho$ and are at least 2, and by Corollary 6.4 and Proposition 6.5, $|\sigma(M) \geqslant \prod_{1 \leqslant j \leqslant \rho} a_j \geqslant 2^{\rho - 1}(n - \rho + 2)$. ∎

A final consequence of Corollary 6.4 is

COROLLARY 6.6. *Let A be a set of cardinality n and rank ρ that has no plateau. Let F be a proper flat, $m=|A/F|=n-|F|$, and $k=\rho(A/F)=\rho-\rho(F)$. Let $j\leqslant k$ be a nonnegative integer. Then $|\sigma(A/F)|\geqslant 3^j 2^{k-j-1}(m-k-j+2)$.*

Proof. Since A has no plateau, we can sequentially construct a sequence of flats $F=F_0, F_1, ..., F_k=A$ such that F_i covers F_{i-1} and $|F_i-F_{i-1}|\geqslant 2$. Let $E_i=F_i/F$. Then $E_0, E_1, ..., E_k=A/F$ is a flag of A/F and so by Corollary 6.4 and Proposition 6.5,

$$|\sigma(A)|\geqslant \prod_{1\leqslant i\leqslant k}(|E_i-E_{i-1}|+1)=\prod_{1\leqslant i\leqslant k}(|F_i-F_{i-1}|+1)$$
$$\geqslant 3^{k-1}(m-2k+3)\geqslant 3^j 2^{k-j-1}(m-k-j+2),$$

where the last inequality follows from $3^a b\geqslant 2^a(a+b)$ for b at least 1 and a nonnegative. ∎

VII. SOME ARITHMETIC INEQUALITIES

The proofs of Theorem 5.2 (which implies Theorem 1.4) and Theorem 5.3 (which implies Theorem 1.2) are given in Sections 8 and 9. The proofs are inductive. Lemma 6.1 and its corollaries are used to reduce the inductive step to various routine but tedious arithmetic inequalities. So as not to disrupt the flow of the proofs, the needed arithmetic lemmas are collected in this section. On first reading, the reader is advised to skim this section and refer to it as needed.

7.1. *Arithmetic Facts about $s(a)$*

The first set of facts we need are about the function $s(a)$, which is equal to $(a^2+a+2)/2$ by Theorem 1.1. The following fact is immediate from the definition.

PROPOSITION 7.1. *For any a and b, $(s(a+b)-s(a))/(s(b)-1)=1+2a/(b+1)$.*

LEMMA 7.2. *If $a\geqslant 3$ and $b\geqslant 2$ then $s(a)\,s(b)\geqslant 4s(a+b-2)$.*

Proof. Let $u=a-3$ and $v=b-2$. Then the desired inequality is equivalent to

$$(u^2+7u+14)(v^2+5v+8)\geqslant 8((u+v)^2+7(u+v)+14).$$

It can be checked that the coefficient of each monomial in u and v on the

left is less than or equal to the coefficient of the same monomial on the right, so the inequality holds for all nonnegative u and v. ∎

LEMMA 7.3. *If a and b satisfy $a \geqslant b \geqslant 1$ and $a \geqslant 5$, then* $(b+2)\, s(a) \geqslant 2s(a+b)$.

Proof. The desired inequality is $(b+2)(a^2+a+2) \geqslant 2((a+b)^2+(a+b)+2)$ which reduces to $a^2 \geqslant 3a+2b$ and follows from the hypotheses. ∎

LEMMA 7.4. *If a and b are positive integers satisfying $b \geqslant 1+a/2$ then* $(9/4)\, s(a) \leqslant s(a+b)$.

Proof. $s(a+b)/s(a) = 1+(2ab+b+b^2)/(a^2+a+2)$ which, under the hypothesis, is at least $1+(2+7a/2+5a^2/4)/(a^2+a+2) = 2+(5a/2+a^2/4)/(a^2+a+2) \geqslant 9/4$ for all $a \geqslant 1$. ∎

The induction hypothesis for the proof of Theorem 1.2 will be formulated in terms of a function related to $s(a)$. Let $L = 181$ and for $a > L$, define $r(a) = \min\{s(a), (9/8)\, s(a-L)\}$.

LEMMA 7.5. *There exists a constant K such that $r(a) = (9/8)\, s(a-L)$ for $a < K$ and $r(a) = s(a)$ if $a \geqslant K$. Furthermore, K is between $17L$ and $18L$.*

Proof. $r(a) = 9/8 s(a-L)$ if and only if $s(a) - 9/8 s(a-L) = (-a^2 + a(18L-1) - 9L^2 + 9L - 2)/16$ is positive. This is a quadratic in a which is positive when $a = L$ and has one root K greater than L, and thus changes sign only when a exceeds that root. It is easy to check that it is positive when $a = 17L$ and negative when $a = 18L$. ∎

LEMMA 7.6. *Let $a \geqslant 3$ and $b \geqslant L+2$. Then $s(a)\, r(b) \geqslant 4r(a+b-2)$.*

Proof. Let K be the constant of Lemma 7.5. If $b \geqslant K$ then the desired inequality is equivalent to Lemma 7.2. Otherwise, $r(b) = (9/8)\, s(b-L)$ and by Lemma 7.2, $s(a)\, r(b) = (9/8)\, s(a)\, s(b-L) \geqslant (9/8)\, 4s(a+b-L-2) \geqslant 4r(a+b-2)$. ∎

LEMMA 7.7. *If a and b are positive integers then $r(a+b+L) - r(a+L) \leqslant (9/8)(s(a+b) - s(a))$.*

Proof. If $(a+L) < K$ then, by Lemma 7.5, $r(a+L) = (9/8)\, s(a)$ and the inequality is equivalent to $r(a+b+L) \leqslant (9/8)\, s(a+b)$, which is immediate from the definition. If $a+L \geqslant K$, then Lemma 7.5 implies $r(a+b+L) - r(a+L) = s(a+b+L) - s(a+L) = (2ab+2bL+b^2+b)/2$. So we want $(2ab+2bL+b^2+b)/2 \leqslant (9/8)(2ab+b^2+b)/2$ or $16L \leqslant 2a+b+1$, which holds since $a+L \geqslant K \geqslant 17L$. ∎

7.2. *Arithmetic Facts about* $t^*(a)$

For reference we give a table of small values for $t^*(a)$:

n	1	2	3	4	5	6	7	8	9	10
$t^*(n)$	2	3	4.5	6.75	9.5	13	17	21	26	31

LEMMA 7.8. *Let a and b be positive integers. Then* $t^*(a)\,t^*(b) \geqslant 2t^*(a+b-1)$.

Proof. Without loss of generality, assume $a \leqslant b$. If $b \leqslant 5$, the inequality holds by inspection of the table. So assume that $b \geqslant 6$ and thus $t^*(b) = t(b) = (b^2 + 2b + P(b) + 4)/4$.

If $a = 1$ then $t^*(a) = 2$ and the inequality holds as an equality.

If $a = 2$, we need that $t^*(a)\,t^*(b) = 3(b^2 + 2b + P(b) + 4)/4$ is at least $2t^*(a+b-1) = 2(b^2 + 4b + 8 - P(b))/4$, which is equivalent to $b^2 - 2b - 4 + 5P(b) \geqslant 0$, which holds for $b \geqslant 6$.

If $a = 3$, we need $t^*(a)\,t^*(b) = 4.5(b^2 + 2b + P(b) + 4)/4$ to be at least $2t^*(a+b-1) = 2(b^2 + 6b + 12 + P(b))/4$, which is equivalent to $2.5b^2 - 3b - 6 + 2.5P(b) \geqslant 0$, which holds for $b \geqslant 6$.

Finally, if $a, b \geqslant 4$, then

$$t^*(a)\,t^*(b) \geqslant (a^2 + 2a)(b^2 + 2b)/16 = (a^2b^2 + 2ab^2 + 2a^2b + 4ab)/16$$
$$\geqslant (2ab + b^2 + a^2 + 8)/2,$$

while

$$2t^*(a+b-1) = 2((a+b-1)^2 + 2(a+b-1) + P(a+b-1) + 4)/4$$
$$\leqslant (2ab + b^2 + a^2 + 4)/2,$$

and the desired inequality follows. ∎

For positive integers a and b, let $T(a, b) = (t^*(a+b) - t^*(a))/(t(b) - 1)$.

LEMMA 7.9. *For positive integers a, b,*

(i) $T(a, b) \leqslant 3(a+1)/4$

(ii) *If* $(a, b) \notin \{(4, 2), (5, 2)\}$ *then* $T(a, b) \leqslant 1 + a/2$.

Proof. By simple manipulation,

$$T(a, b) = 1 + \frac{2a}{b + 2 + P(b)/b} + \frac{4(\varepsilon(a) - \varepsilon(a+b)) + P(a+b) - P(a) - P(b)}{b^2 + 2b + P(b)}.$$

Since $b+2+P(b)/b \geqslant 4$ and $b^2+2b+P(b) \geqslant 4$ for all $b \geqslant 1$, we have

$$T(a, b) \leqslant 1 + a/2 + \varepsilon(a) - \varepsilon(a+b) + (P(a+b) - P(a) - P(b))/4. \quad (7.1)$$

Since $P(a+b) \leqslant P(a) + P(b)$, we have $T(a, b) \leqslant 1 + a/2$, unless $\varepsilon(a) \geqslant \varepsilon(a+b)$ which happens only if $(a, b) \notin \{(3, 1), (4, 2), (5, 1), (5, 2)\}$.

For $(a, b) = (3, 1)$ or $(a, b) = (5, 1)$, $(P(a, b) - P(a) - P(b))/4 = -1/2$ while $\varepsilon(a) - \varepsilon(a+b) \leqslant 1/2$ and thus $T(a, b) \leqslant 1 + a/2$ unless $(a, b) \in \{(4, 2), (5, 2)\}$ proving (ii). To prove (i) note that $1 + a/2 \leqslant 3(a+1)/4$ for all $a \geqslant 1$ and thus from (ii) we have $T(a, b) \leqslant 3(a+1)/4$ unless $a = 4$ or 5. But also, by (7.1), $T(a, b) \leqslant 3/2 + a/2$ (since $\varepsilon(a) \leqslant 1/2$) which for $a = 4$ or 5 is less than $3(a+1)/4$. ∎

VIII. Proofs of Theorems 1.2 and 5.3

The proof of Theorem 1.2 requires a stronger induction hypothesis. Let $L = 181$. For $m > L$, define the function $r(m) = \min\{s(m), (9/8)\, s(m-L)\}$. By Lemma 7.5, there is a number K (approximately $18L$) such that $r(m) = (9/8)\, s(m-L)$ for $m \leqslant K$ and $r(m) = s(m)$ for $m \geqslant K$. Theorem 1.2 then follows from:

Theorem 8.1. *For all* $n \geqslant \rho + L$,

$$s_\rho(n) \geqslant 2^{\rho-1} r(n - \rho + 1).$$

The first step in the proof is a weaker lower bound on $s_\rho(n)$:

Lemma 8.2. *For all* $\rho \leqslant n$, $s_\rho(n) \geqslant 2^{\rho-2} s(n - \rho + 2)$.

Both Lemma 8.2 and Theorem 8.1 are proved by induction on ρ; the basis step $\rho = 1$ is easily verified for both. For the proofs of the induction steps, we assume $\rho > 1$ and let A denote an antipode-free set of cardinality n and rank ρ, and show that $|\sigma(A)|$ is at least the quantity given by the bound.

Proof of Lemma 8.2. If A has a splitting flat F, $x = |F|$, $j = \rho(F)$, $y = n - x = |A/F|$ and $k = \rho - j = \rho(A/F)$. By Corollary 6.3 and the definition of splitting flat,

$$|\sigma(A)| \geqslant (|\sigma(F)|)\, |\sigma(A/F)| \geqslant s_j(x)\, s_k(y).$$

If $x = j$, then by induction, $s_j(x)\, s_k(y) \geqslant 2^j(2^{k-2} s(y - k + 2)) = 2^{\rho-2} s(n - \rho + 2)$. The case $y = k$ is handled similarly. Otherwise, both $x - j$

and $y-k$ are positive. Applying the induction hypothesis and Lemma 7.2 with $a=x-j+2$ and $b=y-k+2$ gives

$$\begin{aligned} s_j(x)\, s_k(y) &\geqslant 2^{j-2}s(x-j+2)\, 2^{k-2}s(y-k+2) \\ &\geqslant 2^{j+k-2}s((x-j+2)+(y-k+2)-2) \\ &\geqslant 2^{\rho-2}s(n-\rho+2), \end{aligned}$$

as required.

Thus, it suffices to consider the case that A does not have a splitting flat. Let F be any rank 1 flat, $x=|F|$ and $y=n-x=|A/F|$. Since F is not a splitting flat, Proposition 2.2(ii) implies $\rho(A-F)=\rho$. By Lemma 6.1, the induction hypothesis, Theorem 1.5, and Proposition 1.1,

$$\begin{aligned} |\sigma(A)| &\geqslant |\sigma(A-F)| + |\sigma(A/F)|\ (|\sigma(F)|-1) \\ &\geqslant 2^{\rho-2}s(y-\rho+2) + 2^{\rho-2}(y-\rho+3)(s(x)-1). \end{aligned}$$

The induction step follows if the right hand side is at least $2^{\rho-2}s(y+x-\rho+2)$. This is equivalent to $(s(y+x-\rho+2)-s(y-\rho+2))/(s(x)-1)\leqslant y-\rho+3$. By Proposition 7.1, with $a=y-\rho+2$ and $b=x$, the left hand side of this inequality is equal to $1+2(y-\rho+2)/(x+1)\leqslant y-\rho+3$, since x is at least 1. ∎

Proof of Theorem 8.1. If $L+1\leqslant n-\rho\leqslant 3L-1$ then $b=L+1$ and $a=n-\rho-L+1$ satisfy the hypotheses of Lemma 7.4 and thus Lemmas 7.4 and 8.2 imply $2^{\rho-1}r(n-\rho+1)\leqslant 2^{\rho-1}(9/8)\, s(n-\rho-L+1)\leqslant 2^{\rho-2}s(n-\rho+2)\leqslant s_\rho(n)$. So assume that $n\geqslant \rho+3L$ and, as before, let A be a set of vectors of cardinality n and rank ρ. If A has a splitting flat F, let $x=|F|$, $j=\rho(F)$, $y=n-x=|A/F|$ and $k=\rho-j=\rho(A/F)$. By Corollary 6.3 and the definition of splitting flat, $|\sigma(A)|\geqslant|\sigma(A/F)|\ (|\sigma(F)|)\geqslant s_j(x)\, s_k(y)$. Assume without loss of generality that $x-j\leqslant y-k$. Then $y-k$ is at least $(n-\rho)/2>L+1$. By induction, Lemma 8.2, and Lemma 7.6,

$$\begin{aligned} |\sigma(A)| &\geqslant s_j(x)\, s_k(y) \geqslant 2^{j-2}s(x-j+2)\, 2^{k-1}r(y-k+1) \\ &\geqslant 2^{j+k-1}r((x-j)+(y-k)+1) = 2^{\rho-1}r(n-\rho+1), \end{aligned}$$

as required.

It remains to consider the case that A does not have any splitting flats (and hence, by Proposition 2.2(i), no plateaus). Let F be any flat of rank 1, $x=|F|$ and $y=n-x=|A/F|$. If $x\geqslant n-\rho-L$ then, since $n\geqslant\rho+3L$, we have $x\geqslant y-\rho+1$. Applying Corollary 6.3, Theorem 1.5, Proposition 1.1, and Lemma 7.3 (with $a=x$ and $b=y-\rho+1$),

$$\begin{aligned} |\sigma(A)| &\geqslant |\sigma(A/F)|\ (|\sigma(F)|) \geqslant 2^{\rho-2}(y-\rho+3)(s(x)) \\ &\geqslant 2^{\rho-1}(s(x+y-\rho+1)) \geqslant 2^{\rho-1}r(n-\rho+1), \end{aligned}$$

proving the desired bound.

So we may assume that for any rank one flat, $|F| < n-\rho-L$, and thus $|A-F| > \rho+L$. Since F is not a splitting flat, Proposition 2.2(ii) implies $\rho(A-F)=\rho$. By the induction hypothesis, $|\sigma(A-F)| \geqslant 2^{\rho-1}r(y-\rho+1)$. By Lemma 6.1 and Proposition 1.1,

$$\begin{aligned}|\sigma(A)| &\geqslant |\sigma(A-F)| + |\sigma(A/F)|\ (|\sigma(F)|-1)\\ &\geqslant 2^{\rho-1}r(y-\rho+1) + |\sigma(A/F)|\ (s(x)-1).\end{aligned}$$

Let $z=y-L-\rho+1$. Then $z \geqslant 1$ and the induction step follows if there is a rank one flat F such that

$$2^{\rho-1}(r(z+L+x)-r(z+L)) \leqslant |\sigma(A/F)|\ (s(x)-1).$$

Note that by Lemma 7.7 and Proposition 7.1,

$$\begin{aligned}&(r(z+L+x)-r(z+L))/(s(x)-1)\\ &\quad\leqslant (9/8)(s(z+x)-s(z))/(s(x)-1) \leqslant (9/8)(1+2z/(x+1)),\end{aligned}$$

so it is enough to show that there is a rank one flat F such that

$$|\sigma(A/F)| \geqslant 9(2^{\rho-4})(1+2z/(x+1)). \tag{8.1}$$

Suppose $\rho \geqslant 4$. Then, noting that $\rho(A/F)=\rho-1$, Corollary 6.6 (with $j=2$, $k=\rho-1$ and $m=y$) implies $|\sigma(A/F)| \geqslant 3^2 2^{\rho-4}(y-\rho) = 9(2^{\rho-4})(z+L-1) \geqslant 9(2^{\rho-4})(z+1)$, which implies (8.1) since $x \geqslant 1$.

Suppose $\rho \leqslant 3$. If $x=|F| \geqslant 4$, then by Theorem 1.5,

$$\begin{aligned}|\sigma(A/F)| &\geqslant 2^{\rho-2}(z+L+2) = 9(2^{\rho-4})(4/9)(z+L+2)\\ &\geqslant 9(2^{\rho-4})(1+2z/(x+1)),\end{aligned}$$

where the last inequality holds for all $x \geqslant 4$ and $z \geqslant 1$.

Finally, we are left with the case that $\rho=2$ or 3, A has no splitting flats and every rank one flat of A has size 1, 2, or 3. In this case, A must have at least $n/3$ flats of rank 1, since the rank one flats partition A. Then by Lemma 2.3(i), A has a flat F such that A/F has an antipode free subset B of size at least $(n-3)/9$. By Corollary 3.3(i),

$$|\sigma(A/F)| \geqslant |\sigma(B)| \geqslant (((n-3)/9)^2 + (n-3)/9 + 2)/2 = (n^2+3n+144)/162.$$

Since $n = x+y = x+\rho+L-1+z \geqslant z+L+2$,

$$|\sigma(A/F)| \geqslant (z^2+2zL+L^2+7z+7L+154)/162.$$

Also since $\rho \leqslant 3$, the right hand side of (8.1) is at most $(9/2)(1+z)$. Thus

(8.1) follows in this case if $(z^2+2zL+L^2+7z+7L+154) \geqslant 729(1+z)$ which is equivalent to

$$z(722-2L-z) \leqslant L^2+7L-575.$$

The left hand side is always at most $(361-L)^2$, and so the inequality holds if $L \geqslant 181$. This completes the proof of this case and the proof of Theorem 8.1. ∎

Remark. The constant K obtained by this proof is 3258. There are several places in the above proof where a more detailed calculation can be used to reduce the value of the constant. We guess that the correct value is not more than about 10, but do not know any ways to show this without a tedious case analysis.

Finally, Theorem 1.2 can be used to show that there are at most finitely many s-critical indices:

Proof of Theorem 5.3. We claim that the only possible s-critical indices (ρ, v) have $v+\rho \leqslant 3K$, which implies that there are only finitely many such indices. Suppose that A is a set of vectors of size $n > 3K$ and rank ρ that minimizes $|\sigma(A)|$. If $n-\rho > K$, then Theorem 8.1 implies that $|\sigma(A)| \geqslant 2^{\rho-1}s(n-\rho+1)$, and so (ρ, n) is not critical. Otherwise, $|A| < 3\rho/2$ and by Lemma 2.3(iii), there is a flat F of size one such that A/F is an antipode free set. Then $2s_\rho(n) \geqslant |\sigma(A)| = 2|\sigma(A/F)| \geqslant 2s_{\rho-1}(n-1)$, which implies that (ρ, n) is not s-critical. ∎

IX. Proofs of Theorems 1.4 and 5.2

As noted in Section 5, Theorem 5.2 implies Theorem 1.4 and thus it is enough to prove Theorem 5.2. The proof is similar to that of Theorem 8.1 in the previous section. We proceed by induction on ρ, and for fixed ρ, by induction on n. The basis step, $\rho = 1$, is Theorem 1.3 (which follows from Corollary 3.3(ii)). For the proof of the induction steps, we assume $\rho > 1$ and let A denote a set of nonzero vectors of cardinality n and rank ρ, and show that $|\sigma(A)| \geqslant t_\rho^*(n)$. It is enough to show $|\sigma(A)| \geqslant 2^{\rho-1}t^*(n-\rho+1)$ since $t_\rho^*(n) = 2^{\rho-1}t^*(n-\rho+1)$ unless $\rho = 2$ and $n = 5$, in which case $|\sigma(A)| \geqslant 2^{\rho-1}t^*(n-\rho+1)$ and $|\sigma(A)|$ an integer imply $\lceil 2^{\rho-1}t^*(n-\rho+1)\rceil = \lceil 13.5\rceil = t_\rho^*(n)$.

If A has a splitting flat F then let $x = |F|$, $j = \rho(F)$, $y = n-x = |A/F|$ and $k = \rho - j = \rho(A/F)$. By Lemma 7.8 and induction, $|\sigma(A)| \geqslant |\sigma(A/F)|\,(|\sigma(F)|) \geqslant t_j^*(x)\, t_k^*(y) \geqslant t_{j+k}^*(x+y) = t_\rho^*(n)$.

So we may assume that A has no splitting flats. Then by Proposition 2.2(ii), $\rho(A-F)=\rho(A)$ for any rank one flat F of A. Thus if F is any rank one flat, we may apply Lemma 6.1, the induction hypothesis and Theorem 1.3 to conclude that

$$\begin{aligned}|\sigma(A)| &\geqslant |\sigma(A-F)| + |\sigma(A/F)|\ (|\sigma(F)|-1) \\ &\geqslant 2^{\rho-1}t^*(n-|F|-\rho+1) + |\sigma(A/F)|\ t(|F|-1). \qquad (9.1)\end{aligned}$$

To complete the proof of the theorem it is enough to show that the right hand side of (9.1) is at least $2^{\rho-1}t^*(n-\rho+1)$ which follows from:

LEMMA 9.1. *Let A be an antipode free set that has no splitting flats, and let $n=|A|$ and $\rho=\rho(A)\geqslant 2$. Then A has a rank one flat F such that*

$$|\sigma(A/F)| \geqslant 2^{\rho-1}T(n-|F|-\rho+1, |F|).$$

(Here, T is the function defined before Lemma 7.9.)

Proof. Case *i.* $\rho\geqslant 3$. Then by Corollary 6.6 and the hypothesis that A has no splitting flats,

$$|\sigma(A/F)| \geqslant 3(2^{\rho-3})(n-|F|-\rho+2)$$

for any rank one flat F. By Lemma 7.9(i), this is at least $2^{\rho-1}T(n-|F|-\rho+1, |F|)$.

Case ii. $\rho=2$. The desired inequality now simplifies to

$$|\sigma(A/F)| \geqslant 2T(n-|F|-1, |F|)$$

for some rank one flat F. Now, by Theorem 1.5, $|\sigma(A/F)|\geqslant n-|F|+1$. By Lemma 7.9(ii), this at least $2T(n-|F|-1, |F|)$ unless $|F|=2$ and $n=7$ or 8. Thus we may assume that $n=7$ or 8 and all rank one flats F of A have exactly two elements. Since A is the disjoint union of its rank one flats, $n=8$.

So now we need that for some rank one flat F, $|\sigma(A/F)|\geqslant \frac{15}{2}$.

CLAIM. *For one of the rank one flats of A, A/F has more than one distinct antipode class.*

Assuming the claim, we have by Theorem 3.1 that $|\sigma(A/F)|\geqslant 2+|A/F|=8$ completing the proof of the lemma. So it remains to prove the claim. Suppose the claim is false, i.e., that for each rank one flat F, A/F consists of one antipode class. If some rank one flat E does not consist of antipodal vectors then if we take F to be any other rank one flat, the two elements of E are in different antipode classes in A/F. Thus each of the four

rank one flats of F consists of an antipodal pair of vectors. Let $\mathbf{w}$, $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ be representatives of the rank one flats. Since $\mathbf{w}$ and $\mathbf{x}$ are in the same antipode class in $A/\langle \mathbf{y} \rangle$, then we have that either $\mathbf{w}+\mathbf{x}$ or $\mathbf{w}-\mathbf{x}$ is equal to $a\mathbf{y}$ for some constant a. Similarly, either $\mathbf{w}+\mathbf{y}$ or $\mathbf{w}-\mathbf{y}$ is equal to $b\mathbf{x}$ for some constant b. Now the representation of $\mathbf{w}$ in the form $c\mathbf{x}+d\mathbf{y}$ is unique, which from the above implies that $c, d \in \{-1, +1\}$. A symmetric argument, says that $\mathbf{w} = e\mathbf{x} + f\mathbf{z}$ and $\mathbf{w} = g\mathbf{y} + h\mathbf{z}$ where $e, f, g, h \in \{-1, 1\}$. Now we must have either $c = e$, $d = g$ or $h = f$, since otherwise summing these equations gives $\mathbf{w} = \mathbf{0}$, a contradiction. Consider the case $c = e$ (the other two cases are similar). Then $d\mathbf{y} = \mathbf{w} - c\mathbf{x} = \mathbf{w} - e\mathbf{x} = f\mathbf{z}$, which contradicts the fact that $\mathbf{y}$ and $\mathbf{z}$ are in distinct rank one flats. This contradiction proves the claim, thus completing the proofs of Lemma 9.1 and Theorems 1.4 and 5.2.

X. Proof of Theorem 1.6

Let C be a subset of $\mathbf{R}$. A map g from C to $\mathbf{R}^k$ is said to be *spanning* if the image $g(C)$ spans $\mathbf{R}^k$. The map g is σ-preserving if for any sequence $(\varepsilon_c : c \in C)$ with $\varepsilon_c \in \{-1, 0, 1\}$,

$$\sum_{c \in C} \varepsilon_c c = 0 \text{ implies } \sum_{c \in C} \varepsilon_c g(c) = \mathbf{0},$$

and it is σ-injective if for any sequence $(\varepsilon_c : c \in C)$ with $\varepsilon_c \in \{-1, 0, 1\}$,

$$\sum_{c \in C} \varepsilon_c c \neq 0 \text{ implies } \sum_{c \in C} \varepsilon_c g(c) \neq \mathbf{0}.$$

In particular, note that a σ-injective map g must be one-to-one and $g(c) \neq -g(d)$ for two distinct elements of C. Note that if g is σ-preserving and σ-injective then $|\sigma(C)| = |\sigma(g(C))|$.

Theorem 1.6 follows from

Theorem 10.1. *Let a and b be positive integers and $n = a + b$. Let $B = \{c_1, c_2, ..., c_b\}$ be a set of positive integers of size b such that $c_k / c_{k-1} > 2n^{n-1}$ for each k between 2 and b. Let A be any set of a positive integers disjoint from B. Then there is a map g from $A \cup B$ to R^b such that g is spanning, σ-preserving, and σ-injective.*

To deduce Theorem 1.6 from this, suppose that MAX chooses a set B that satisfies the hypotheses of Theorem 10.1 and that MIN chooses any set A disjoint from B. Then if g is the function whose existence is guaranteed

by Theorem 10.1, then $|\sigma(A \cup B)| = |\sigma(g(A \cup B))| \geqslant s_b(a+b)$ since $g(A \cup B)$ is an antipode-free set of vectors of dimension b.

So it suffices to construct the map g whose existence is asserted by Theorem 10.1. For a real number a, $[a]$ denotes the closest integer to a, i.e., $\lceil a - 1/2 \rceil$. To each sequence $\mathbf{x} = (x_1, x_2, ..., x_k)$ of positive reals we associate a map $a \to a\langle x_1, x_2, ..., x_k \rangle$ from $\mathbf{R}$ to $\mathbf{R}$, given by

$$a\langle\ \rangle = a$$

$$a\langle x_1 \rangle = a - x_1[a/x_1]$$

$$a\langle x_1, x_2, ..., x_k \rangle = a\langle x_1, x_2, ..., x_{k-1} \rangle \langle x_k \rangle,$$

and a map $g^{\mathbf{x}}: \mathbf{R} \to \mathbf{R}^{k+1}$ given by

$$g^{\mathbf{x}}(a) = (a, a\langle x_1 \rangle, a\langle x_1, x_2 \rangle, ..., a\langle x_1, x_2, ..., x_k \rangle).$$

For instance, if $\mathbf{x} = (28, 9, 1.5)$ and $a = 200$ then $g^{\mathbf{x}}(a) = (200, 4, 4, -0.5)$.

Trivially, $g^{\mathbf{x}}$ is a σ-injective map for any $\mathbf{x}$. Given $A \cup B$ as in the theorem, let $B = \{c_1, c_2, ..., c_b\}$ and $A = \{c_{b+1}, c_{b+2}, ..., c_n\}$. We construct a positive vector $\mathbf{x} \in \mathbf{R}^{b-1}$ that satisfies the following two conditions for each $1 \leqslant i \leqslant b-1$:

$P(i)$ For $1 \leqslant j \leqslant b$,

$$c_j\langle x_1, x_2, ..., x_i \rangle = \begin{cases} c_j & \text{if } i < j \\ 0 & \text{if } i \geqslant j \end{cases};$$

$Q(i)$ For $1 \leqslant j \leqslant n$,

$$|c_j\langle x_1, x_2, ..., x_i \rangle| < x_i/n.$$

This suffices to prove Theorem 10.1 by the following:

LEMMA 10.2. *If $\mathbf{x} \in \mathbf{R}^{b-1}$ satisfies $P(i)$ and $Q(i)$ for $1 \leqslant i \leqslant b-1$ with respect to $\mathbf{C} = \{c_1, c_2, ..., c_n\}$ then $g^{\mathbf{x}}$ is spanning and σ-preserving on $\mathbf{C}$.*

Proof. Conditions $\{A(i) | 1 \leqslant i \leqslant b-1\}$ easily imply that imply that $g^{\mathbf{x}}(c_1), g^{\mathbf{x}}(c_2), ..., g^{\mathbf{x}}(c_b)\}$ spans $\mathbf{R}^b$. To see that $g^{\mathbf{x}}$ is σ-preserving on $\mathbf{C}$, suppose that $(\varepsilon_j: 1 \leqslant j \leqslant n) \in \{-1, 0, 1\}^n$ with $\sum_{j=1}^n \varepsilon_j c_j = 0$. We show by induction on i that $\sum_{j=1}^n \varepsilon_j c_j \langle x_1, x_2, ..., x_i \rangle = 0$. For $i = 0$ this is the hypothesis. Let $i \geqslant 1$. Defining $d_j = c_j\langle x_1, x_2, ..., x_{i-1} \rangle$, the induction hypothesis is $\sum_{j=1}^n \varepsilon_j d_j = 0$. Thus

$$\sum_{j=1}^{n} \varepsilon_j c_j \langle x_1, x_2, ..., x_i \rangle = \sum_{j=1}^{n} \varepsilon_j (d_j - [d_j/x_i]\, x_i) = -\left(\sum_{j=1}^{n} \varepsilon_j [d_j/x_i]\, x_i \right),$$

which is an integer multiple of x_i and is therefore $\mathbf{0}$ since by condition $Q(i)$,

$$\left|\sum_{j=1}^{n} \varepsilon_j c_j \langle x_1, x_2, ..., x_i\rangle\right| \leqslant \sum_{j=1}^{n} |c_j\langle x_1, x_2, ..., x_i\rangle| < \sum_{i=1}^{n} x_i/n = x_i.$$

This completes the proof of the lemma. ▮

The sequence $x_1, x_2, ..., x_{b-1}$ is constructed inductively as follows. We assume that $x_1, x_2, ..., x_{i-1}$ have been chosen to satisfy $\{P(h): 1 \leqslant h \leqslant i-1\}$ and $\{Q(h): 1 \leqslant h \leqslant i-1\}$ and choose x_i so that $P(i)$ and $Q(i)$ are both satisfied. For $1 \leqslant j \leqslant n$, let $d_j = c_j\langle x_1, x_2, ..., x_{i-1}\rangle$ as above.

$P(i-1)$ implies that for $1 \leqslant j \leqslant b$

$$d_j = \begin{cases} 0 & \text{if } j < i \\ c_j & \text{if } j \geqslant i \end{cases}.$$

Let $r_j = d_j/d_i = d_j/c_i$ for $1 \leqslant j \leqslant n$.

CLAIM. *If m is an integer such that*

$$1 \leqslant m \leqslant n^{n-1} \tag{10.1}$$

and

$$|r_j m - [r_j m]| < 1/n \quad \text{for} \quad 1 \leqslant j \leqslant n \tag{10.2}$$

then $x_i = c_i/m$ satisfies $P(i)$ and $Q(i)$.

Proof of Claim. It is easy to check that (10.2) is simply a reformulation of $Q(i)$. Now for $P(i)$ to be satisfied we need

$$d_j - [d_j/x_i]\, x_i = \begin{cases} 0 & \text{if } j \leqslant i \\ c_j & \text{if } i < j \leqslant b \end{cases}.$$

For $j \leqslant i-1$, $d_j = 0$ so this is trivial. For $j = i$, $d_j/x_i = m$ is an integer so $d_j - [d_j/x_i]\, x_i = 0$. For $i < j \leqslant b$, $d_j - [d_j/x_i]\, x_i = c_j - [c_j m/c_i]\, c_i/m = c_j$ since $|c_j m/c_i| < |m/2n^{n-1}| \leqslant 1/2$ implies $[c_j m/c_i] = 0$.

So it suffices to show the existence of m satisfying (10.1) and (10.2). This is obtained by the Dirichlet pigeon-hole argument [2, p. 156], as follows.

Partition $[0, 1)$ into n intervals $\{I_h: 0 \leqslant h \leqslant n-1\}$, where $I_h = [h/n, (h+1)/n)$. For any real number y, let $I(y)$ be the interval I_h containing $y - \lfloor y \rfloor$. For any integer k, let $\Lambda(k)$ be the sequence $(I(kr_1), I(kr_2), ..., I(kr_n))$. Note that $I(kr_i) = I_0$ so there are at most n^{n-1} possible values for $\Lambda(k)$. By the pigeonhole principle, there exist integers $m_1 < m_2$ in

the range $[1, 1+n^{n-1}]$ such that $A(m_1)=A(m_2)$. We claim $m=m_2-m_1$ satisfies (10.1) and (10.2). Trivially $m \leqslant n^{n-1}$. Now we have for $1 \leqslant j \leqslant n$

$$|(m_2 r_j - \lfloor m_2 r_j \rfloor) - (m_1 r_j - \lfloor m_1 r_j \rfloor)| < 1/n.$$

Letting $p = \lfloor m_2 r_j \rfloor - \lfloor m_1 r_j \rfloor$ we have

$$|mr_j - p| < 1/n$$

and thus $p = \lfloor mr_j \rfloor$, and (10.2) holds.

This completes the construction of $x_1, x_2, \ldots, x_{b-1}$ satisfying $P(i)$ and $Q(i)$ for $1 \leqslant i \leqslant b-1$ which, with Lemma 10.2, proves Theorem 10.1.

XI. Open Problems

One obvious question is to determine all of the s-critical indices and thus exactly determine the function $s_\rho(n)$. We believe that there probably are not very many, but we do not know any reasonable way (short of enumerating many cases) to show this.

The results of Theorems 1.2, 1.4, and 1.5 give the minimum of $|\sigma(A)|$ over sets (multisets) of vectors A of size n and rank d subject to certain restrictions. The problem can be investigated under other natural restrictions on A. A set of vectors is *multiple-free* if it does not contain two vectors $\mathbf{v}$ and $\mathbf{w}$ that are scalar multiples of each other.

Question 11.1. What is the minimum of $|\sigma(A)|$ over multiple-free subsets A of size n and rank ρ?

Another problem is to extend Theorem 3.1 to higher dimensional cases:

Question 11.2. Let $\rho \leqslant t$ be positive integers and $m_1, m_2, \ldots, m_t$ be a sequence of positive integers. Suppose that M is a multiset with rank ρ and support of size t, such that the ith vector has multiplicity m_i. What is the smallest that $|\sigma(M)|$ can be?

Analogous questions can be posed for subsets of finite Abelian groups.

Acknowledgments

Thanks to Joel Spencer for telling us the problem about the MAX–MIN game, to Mark Purtil for helpful conversations, and to Bill Burley for checking over the manuscript.

References

1. B. Bollabas, "Extremal Graph Theory," Academic Press, New York, 1978.
2. G. Hardy and E. M. Wright, "An Introduction to the Theory of Numbers," Oxford Science Publications, London/New York, 1979.